

1 **Common brain disorders are associated with heritable patterns of apparent aging of the**  
2 **brain**

3 Tobias Kaufmann<sup>1,\*</sup>, Dennis van der Meer<sup>1</sup>, Nhat Trung Doan<sup>1</sup>, Emanuel Schwarz<sup>2</sup>, Martina J.  
4 Lund<sup>1</sup>, Ingrid Agartz<sup>1,3,4</sup>, Dag Alnæs<sup>1</sup>, Deanna M. Barch<sup>5,6,7</sup>, Ramona Baur-Streubel<sup>8</sup>, Alessandro  
5 Bertolino<sup>9,10</sup>, Francesco Bettella<sup>1</sup>, Mona K. Beyer<sup>11,12</sup>, Erlend Bøen<sup>3</sup>, Stefan Borgwardt<sup>13,14</sup>,  
6 Christine L. Brandt<sup>1</sup>, Jan Buitelaar<sup>15,16</sup>, Elisabeth G. Celius<sup>17,18</sup>, Simon Cervenka<sup>4</sup>, Annette  
7 Conzelmann<sup>19</sup>, Aldo Córdova-Palomera<sup>1,20</sup>, Anders M. Dale<sup>21</sup>, Dominique J.-F. de Quervain<sup>22,23</sup>,  
8 Pasquale Di Carlo<sup>10</sup>, Srdjan Djurovic<sup>24,25</sup>, Erlend S. Dørum<sup>1,26,27</sup>, Sarah Eisenacher<sup>2</sup>, Torbjørn  
9 Elvsåshagen<sup>1,11,18</sup>, Thomas Espeseth<sup>26</sup>, Helena Fatouros-Bergman<sup>4</sup>, Lena Flyckt<sup>4</sup>, Barbara  
10 Franke<sup>28</sup>, Oleksandr Frei<sup>1</sup>, Beathe Haatveit<sup>1,26</sup>, Asta K. Håberg<sup>29,30</sup>, Hanne F. Harbo<sup>18,11</sup>,  
11 Catharina A. Hartman<sup>31</sup>, Dirk Heslenfeld<sup>32,33</sup>, Pieter J. Hoekstra<sup>34</sup>, Einar A. Høgestøl<sup>11,18</sup>, Terry  
12 Jernigan<sup>35,36,37</sup>, Rune Jonassen<sup>38</sup>, Erik G. Jönsson<sup>1,4</sup>, Karolinska Schizophrenia Project (KaSP)<sup>39</sup>,  
13 Peter Kirsch<sup>40,41</sup>, Iwona Kłoszewska<sup>42</sup>, Knut-Kristian Kolskår<sup>1,26,27</sup>, Nils Inge Landrø<sup>26,3</sup>,  
14 Stephanie Le Hellard<sup>25</sup>, Klaus-Peter Lesch<sup>43,44,45</sup>, Simon Lovestone<sup>46</sup>, Arvid Lundervold<sup>47,48</sup>,  
15 Astri J. Lundervold<sup>49,50</sup>, Luigi A. Maglanoc<sup>1,26</sup>, Ulrik F. Malt<sup>11,51</sup>, Patrizia Mecocci<sup>52</sup>, Ingrid  
16 Melle<sup>1</sup>, Andreas Meyer-Lindenberg<sup>2</sup>, Torgeir Moberget<sup>1</sup>, Linn B. Norbom<sup>1,26</sup>, Jan Egil Nordvik<sup>53</sup>,  
17 Lars Nyberg<sup>54</sup>, Jaap Oosterlaan<sup>32,55,56</sup>, Marco Papalino<sup>10</sup>, Andreas Papassotiropoulos<sup>57,23,58</sup>, Paul  
18 Pauli<sup>8</sup>, Giulio Pergola<sup>10</sup>, Karin Persson<sup>59,60</sup>, Geneviève Richard<sup>1,26,27</sup>, Jaroslav Rokicki<sup>1,26</sup>, Anne-  
19 Marthe Sanders<sup>1,26,27</sup>, Geir Selbæk<sup>60,61,17</sup>, Alexey A. Shadrin<sup>1</sup>, Olav B. Smeland<sup>1</sup>, Hilkka  
20 Soininen<sup>62,63</sup>, Piotr Sowa<sup>12</sup>, Vidar M. Steen<sup>25,64</sup>, Magda Tsolaki<sup>65</sup>, Kristine M. Ulrichsen<sup>1,26,27</sup>,  
21 Bruno Vellas<sup>66</sup>, Lei Wang<sup>67</sup>, Eric Westman<sup>68,14</sup>, Georg C. Ziegler<sup>43</sup>, Mathias Zink<sup>2,69</sup>, Ole A.  
22 Andreassen<sup>1</sup>, Lars T. Westlye<sup>1,26,\*</sup>

23 1 NORMENT, Division of Mental Health and Addiction, Oslo University Hospital & Institute of  
24 Clinical Medicine, University of Oslo, Oslo, Norway

25 2 Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty  
26 Mannheim, Heidelberg University, Mannheim, Germany

27 3 Department of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway

28 4 Department of Clinical Neuroscience, Center for Psychiatry Research, Karolinska Institutet and  
29 Stockholm County Council, Stockholm, Sweden

30 5 Department of Psychological and Brain Sciences, Washington University in St. Louis, St. Louis, USA

31 6 Department of Psychiatry, Washington University in St. Louis, St. Louis, USA

32 7 Department of Radiology, Washington University in St. Louis, St. Louis, USA

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

- 33 8 Department of Psychology I, University of Würzburg, Würzburg, Germany  
34 9 Institute of Psychiatry, Bari University Hospital, Bari, Italy  
35 10 Department of Basic Medical Science, Neuroscience, and Sense Organs, University of Bari, Bari, Italy.  
36 11 Institute of Clinical Medicine, University of Oslo, Oslo, Norway  
37 12 Department of Radiology and Nuclear Medicine, Section of Neuroradiology, Oslo University Hospital,  
38 Oslo, Norway  
39 13 Department of Psychiatry (UPK), University of Basel, Basel, Switzerland  
40 14 Institute of Psychiatry, King's College, London, UK  
41 15 Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour,  
42 Radboud University Medical Center, Nijmegen, The Netherlands  
43 16 Karakter Child and Adolescent Psychiatry University Centre, Nijmegen, The Netherlands  
44 17 Institute of Health and Society, University of Oslo, Oslo, Norway  
45 18 Department of Neurology, Oslo University Hospital, Oslo, Norway  
46 19 Children and Adolescence Psychiatry, University of Tübingen, Tübingen, Germany  
47 20 Department of Pediatrics, Stanford University School of Medicine, Stanford University, Stanford, USA  
48 21 Department of Radiology, University of California, San Diego, La Jolla, CA, USA; Department of  
49 Neurosciences, University of California, San Diego, La Jolla, CA, USA  
50 22 Division of Cognitive Neuroscience, University of Basel, Basel, Switzerland  
51 23 Transfaculty Research Platform Molecular and Cognitive Neurosciences, University of Basel, Basel,  
52 Switzerland  
53 24 Department of Medical Genetics, Oslo University Hospital, Oslo, Norway  
54 25 NORMENT, Department of Clinical Science, University of Bergen, Bergen, Norway  
55 26 Department of Psychology, University of Oslo, Oslo, Norway  
56 27 Sunnaas Rehabilitation Hospital HT, Nesodden, Norway  
57 28 Departments of Human Genetics and Psychiatry, Donders Institute for Brain, Cognition and  
58 Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands  
59 29 Department of Neuromedicine and Movement Science, Norwegian University of Science and  
60 Technology, Trondheim, Norway  
61 30 Department of Radiology and Nuclear Medicine, St. Olavs Hospital, Trondheim, Norway  
62 31 Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen,  
63 The Netherlands  
64 32 Department of Clinical Neuropsychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands  
65 33 Department of Cognitive Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands  
66 34 Department of Child and Adolescent Psychiatry, University Medical Center Groningen, University of  
67 Groningen, Groningen, The Netherlands

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

- 68 35 Center for Human Development, University of California, San Diego, USA  
69 36 Department of Cognitive Science, University of California, San Diego, USA  
70 37 Departments of Psychiatry and Radiology, University of California, San Diego, USA  
71 38 Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway  
72 39 A full list of authors can be found at the end of the article  
73 40 Department of Clinical Psychology, Central Institute of Mental Health, Medical Faculty Mannheim,  
74 Heidelberg University, Mannheim, Germany  
75 41 Bernstein Center for Computational Neuroscience Heidelberg/Mannheim, Mannheim, Germany  
76 42 Department of Old Age Psychiatry and Psychotic Disorders, Medical University of Lodz, Poland  
77 43 Division of Molecular Psychiatry, Center of Mental Health, University of Würzburg, Würzburg,  
78 Germany  
79 44 Laboratory of Psychiatric Neurobiology, Institute of Molecular Medicine, Sechenov First Moscow  
80 State Medical University, Moscow, Russia  
81 45 Department of Neuroscience, School for Mental Health and Neuroscience (MHeNS), Maastricht  
82 University, Maastricht, The Netherlands  
83 46 Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, UK  
84 47 Department of Biomedicine, University of Bergen, Norway  
85 48 Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University  
86 Hospital, Bergen, Norway  
87 49 Department of Biological and Medical Psychology, University of Bergen, Norway  
88 50 K. G. Jebsen Centre for Neuropsychiatric Disorders, University of Bergen, Norway  
89 51 Department of Research and Education, Oslo University Hospital, Oslo, Norway  
90 52 Institute of Gerontology and Geriatrics, University of Perugia, Perugia, Italy  
91 53 CatoSenteret Rehabilitation Center, Son, Norway  
92 54 Department of Radiation Science, Umeå University, Umeå, Sweden  
93 55 Emma Children's Hospital Amsterdam Medical Center, Amsterdam, The Netherlands  
94 56 VU University Medical Center, Department of Pediatrics, Amsterdam, The Netherlands  
95 57 Division of Molecular Neuroscience, University of Basel, Basel, Switzerland  
96 58 Life Sciences Training Facility, Department Biozentrum, University of Basel, Basel, Switzerland  
97 59 Department of Geriatric Medicine, Oslo University Hospital, Oslo, Norway  
98 60 Norwegian National Advisory Unit on Ageing and Health, Vestfold Hospital Trust, Tønsberg, Norway  
99 61 Centre for Old Age Psychiatric Research, Innlandet Hospital Trust, Ottestad, Norway  
100 62 Department of Neurology, Institute of Clinical Medicine, University of Eastern Finland, Kuopio,  
101 Finland

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

- 102 63 Neurocenter, Neurology, Kuopio University Hospital, Kuopio, Finland  
103 64 Dr. E. Martens Research Group for Biological Psychiatry, Department of Medical Genetics, Haukeland  
104 University Hospital, Bergen, Norway.  
105 65 1<sup>st</sup> Department of Neurology, Aristotle University of Thessaloniki, Makedonia, Greece.  
106 66 INSERM U 1027, University of Toulouse, Toulouse, France  
107 67 Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA  
108 68 Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden  
109 69 District hospital Ansbach, Germany

110

111 \* Corresponding authors:

- 112 Tobias Kaufmann, Ph.D. & Lars T. Westlye, Ph.D.  
113 Email: tobias.kaufmann@medisin.uio.no, l.t.westlye@psykologi.uio.no  
114 Postal address: OUS, PoBox 4956 Nydalen, 0424 Oslo, Norway  
115 Telephone: +47 23 02 73 50, Fax: +47 23 02 73 33

116

117 Counts:

- 118 Abstract: 67 words  
119 Main text body: 2880 words  
120 References: 20 in the main paper  
121 Figures: 3 (3 x 2-column)

122

- 123 Key words: Brain age gap, brain disorders, genetic architecture, pleiotropy

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

124 **Common risk factors for psychiatric and other brain disorders likely converge on biological**  
125 **pathways influencing the development and maintenance of brain structure and function**  
126 **across life. Using structural magnetic resonance imaging data from 45,615 individuals aged**  
127 **3 to 96 years, we demonstrate distinct patterns of apparent brain aging in several brain**  
128 **disorders and reveal genetic pleiotropy between apparent brain aging in healthy individuals**  
129 **and common brain disorders.**

130 Psychiatric disorders and other brain disorders are among the main contributors to morbidity and  
131 disability around the world<sup>1</sup>. The disease mechanisms are complex, spanning a wide range of  
132 genetic and environmental contributing factors<sup>2</sup>. The inter-individual variability is large, but on a  
133 group-level, patients with common brain disorders perform worse on cognitive tests, are less  
134 likely to excel professionally, and engage in adverse health behaviours more frequently<sup>3</sup>. It is  
135 unclear to what extent these characteristics are a cause, consequence or confounder of disease.

136 Dynamic processes influencing the rate of brain maturation and change throughout the  
137 lifespan play a critical role, as reflected in the wide range of disease onset times from early  
138 childhood to old age<sup>4</sup>. This suggests that the age at which individual trajectories diverge from the  
139 norm reflects key characteristics of the underlying pathophysiology. Whereas autism spectrum  
140 disorder (ASD) and attention-deficit/hyperactivity disorder (ADHD) emerge in childhood<sup>5</sup>,  
141 schizophrenia (SZ) and bipolar (BD) spectrum disorders likely develop during late childhood and  
142 adolescence, before the characteristic outbreak of severe symptoms in early adulthood<sup>6</sup>.  
143 Likewise, multiple sclerosis (MS) most often manifests in early adulthood but the disease process  
144 likely starts much earlier<sup>7</sup>. First episodes in major depressive disorder (MDD) can appear at any  
145 stage from adolescence to old age<sup>5</sup>, whereas mild cognitive impairment (MCI) and dementia  
146 (DEM) primarily emerge during senescence<sup>8</sup>. Beyond such differential temporal evolution across

147 the lifespan, age-related deviations from the norm may also differ between disorders in terms of  
148 anatomical location, direction, change rate and magnitude, all of which add complexity to the  
149 interpretation of observed effects.

150 Machine learning techniques enable robust estimation of the biological age of the brain  
151 using information provided by magnetic resonance imaging (MRI)<sup>9,10</sup>, assessing the similarity of  
152 a given brain scan with scans of a range of individuals to estimate the age of the tissue from a  
153 normative lifespan trajectory. Initial evidence suggested that the deviation between brain age and  
154 chronological age – termed the *brain age gap* - is a promising marker of brain health<sup>11</sup>, but  
155 several issues remain to be addressed. First, while advantageous for narrowing the complexity,  
156 reducing a rich set of brain imaging features into a single estimate of brain age inevitably  
157 compromises spatial specificity, thereby neglecting disorder-specific patterns. Second, most  
158 studies so far have been rather small-scale, performed within a limited age range and focusing on  
159 a single disorder, which left them unable to uncover clinical specificity and lifespan dynamics.  
160 Third, the genetic underpinnings of brain age gap are not understood, and it is unknown to what  
161 degree they overlap with the genetic architecture of major clinical traits. To address these critical  
162 knowledge gaps, large imaging genetics samples covering a range of prevalent brain disorders are  
163 necessary.

164 Here, we employed a centralized and harmonized processing protocol including  
165 automated surface-based morphometry and subcortical segmentation using Freesurfer on raw  
166 structural MRI data from 45,615 individuals aged 3 to 96 years that passed quality control  
167 (**Suppl. Fig. 1**). The sample included data from healthy controls (HC;  $n = 39,827$ ; 3-95 years)  
168 and 5,788 individuals with various brain disorders. We included data from individuals with ASD  
169 ( $n = 925$ ; 5-64 years), ADHD ( $n = 725$ ; 7-62 years), prodromal SZ or at risk mental state  
170 (SZRISK;  $n = 94$ ; 16-42 years), SZ ( $n = 1110$ ; 18-66 years), a heterogeneous group with mixed

171 diagnoses in the psychosis spectrum (PSYMIX;  $n = 300$ ; 18-69 years), BD ( $n = 459$ ; 18-66  
172 years), MS ( $n = 254$ ; 19-68 years), MDD ( $n = 208$ ; 18-71 years), MCI ( $n = 974$ ; 38-91 years), and  
173 DEM (including Alzheimer's disease;  $n = 739$ ; 53-96 years). **Suppl. Tables 1-3** provide details  
174 on the sample's characteristics and scanning protocols.

175 We used machine learning to estimate individual brain age based on structural brain  
176 imaging features. First, we grouped all subjects into different samples. For each of the ten clinical  
177 groups, we identified a group of healthy individuals of equal size, matched on age, sex and  
178 scanning site from a pool of 4353 healthy control subjects. All remaining individuals were joined  
179 into one independent sample comprising healthy individuals only. The latter constituted a  
180 training sample, used to train and tune the machine learning models for age prediction ( $n =$   
181 35,474 aged 3-89 years; 18,990 females), whereas the ten clinical samples were used as  
182 independent test samples. **Figure 1a** illustrates the respective age distributions per sex and  
183 diagnosis.

184 The large sample size and wide age-span of the training sample allowed us to model male  
185 and female brain age separately, thereby accounting for potential sexual dimorphisms in brain  
186 structural lifespan trajectories<sup>12</sup>. For each sex, we built a machine learning model based on  
187 gradient tree boosting to predict the age of the brain from a set of thickness, area and volume  
188 features extracted using a multi-modal parcellation of the cerebral cortex as well as a set of  
189 cerebellar/subcortical volume features (1,118 features in total, **Fig. 1b**). Five-fold cross-  
190 validations revealed high correlations between chronological age and predicted brain age ( $r=.93$   
191 and  $r=.94$  for the female and male model, respectively; **Suppl. Fig. 2**). **Suppl. Fig. 3-6** provide  
192 further validation of the prediction approach and **Suppl. Table 4** provides details on sex  
193 differences in the prediction models. Next, we applied the models to predict age for each  
194 individual in the ten independent test samples (predicting brain age using the female model in

195 females and the male model in males) and tested for effects of diagnosis on the brain age gap  
196 using linear models. We used mega-analysis (across-site analysis) as the main statistical  
197 framework and provide results from a meta-analysis framework in the supplement. We included  
198 age, age , sex, scanning site and a proxy of image quality (Euler number) in all statistical models  
199 testing for group differences and clinical associations. To further minimize confounding effects  
200 of data quality, we repeated the main analyses using a more stringent quality control and  
201 exclusion procedure.

202 **Figure 2a** illustrates that the estimated brain age gap was increased in several brain  
203 disorders. Strongest effects were observed in SZ (Cohen's  $d = 0.56$ ), MS ( $d = 0.69$ ), MCI ( $d =$   
204  $0.41$ ) and DEM ( $d = 1.02$ ). PSYMIX ( $d = 0.21$ ) and BD ( $d = 0.27$ ) showed small effects of  
205 increased brain age gap, whereas other groups showed negligible effects ( $d < 0.2$ ). The meta-  
206 analysis converged on the same findings (**Suppl. Fig. 7**) and the results replicated regardless of  
207 the quality control exclusion criterion applied (**Suppl. Fig. 8**). The brain age gap in all clinical  
208 groups was positive on average and there were no signs of a negative brain age gap  
209 (developmental delay) in children with ASD or ADHD, and no significant group by age  
210 interaction effect (**Suppl. Table 5**).

211 We assessed specificity of the spatial brain age gap patterns across clinical groups. We  
212 trained age prediction models using only occipital, frontal, temporal, parietal, cingulate, insula, or  
213 cerebellar/subcortical features (**Fig. 1b**). Cross-validation confirmed the predictive performance  
214 of all regional models (**Suppl. Fig. 2**) which were used to predict regional brain age in the ten  
215 independent test sets. Regional brain age gaps largely corresponded to the full brain level, with  
216 some notable differential spatial patterns (**Fig. 2b**). For example, increased cerebellar/subcortical  
217 age gap was most prominent in DEM ( $d = 0.91$ ) and MS ( $d = 0.82$ ) but was not present in SZ ( $d$   
218  $= 0.10$ ). The largest effect in SZ was observed in the frontal lobe ( $d = 0.72$ ). A brain age gap in



219 the temporal lobe was observed in MDD ( $d = 0.28$ ), whereas there was no evidence ( $d < 0.2$ ) for a  
220 brain age gap in ASD, ADHD or SZRISK in any of the regions. To explore regional differences  
221 in brain age patterns, we tested for group by region interactions on each pairwise combination of  
222 clinical groups and pairwise combination of regional brain age gaps (1260 tests). **Figure 2c**  
223 illustrates the significant effect sizes, indicating that the rate at which different regions age in  
224 relation to each other oftentimes showed opposite patterns between disorders typically considered  
225 neurodevelopmental (e.g. SZ) and neurodegenerative (e.g. MS/DEM), respectively.

226 With converging evidence demonstrating largest brain age gaps in SZ, MS, MCI and  
227 DEM, we explored the functional relevance of the regional brain age gaps for these groups by  
228 testing for associations with clinical and cognitive data. Clinical data available from individuals  
229 with SZ included symptom ( $n = 389$ ) and function ( $n = 269$ ) scores of the Global Assessment of  
230 Functioning scale (GAF) as well as positive ( $n = 646$ ) and negative ( $n = 626$ ) scores of the  
231 Positive and Negative Syndrome Scale (PANSS). For MS, we assessed associations with scores  
232 from the Expanded Disability Status Scale (EDSS,  $n = 195$ ). In the dementia spectrum, we  
233 assessed associations with Mini Mental State Examination scores (MMSE,  $n = 907$  MCI,  $n = 686$   
234 DEM). **Figure 2d** depicts association strengths accounted for age, age, sex, scanning site and  
235 Euler number and **Suppl. Fig. 11** provides corresponding scatter plots. In SZ, larger brain age  
236 gaps were associated with lower functioning, for example full brain age gap with GAF symptom  
237 ( $r = -0.17$ ,  $P = 9 \times 10^{-4}$ ) and insula brain age gap with GAF function ( $r = -0.22$ ,  $P = 3 \times 10^{-4}$ ), and  
238 with more negative symptoms, for example temporal brain age gap with PANSS negative ( $r =$   
239  $0.11$ ,  $P = .005$ ). In MS, larger full brain age gap was associated with higher disability ( $r = 0.24$ ,  $P$   
240  $= .001$ ). Finally, lower cognitive functioning was associated with larger brain age gaps in  
241 MCI/DEM, with strongest effects for full brain ( $r = -0.29$ ,  $P = 2 \times 10^{-29}$ ) and  
242 cerebellar/subcortical ( $r = -0.27$ ,  $P = 1 \times 10^{-26}$ ) brain age gaps.

243           Given the substantial genetic contributions to most brain disorders, our results incite the  
244 question to what degree brain age patterns are genetically influenced and if the implicated  
245 polymorphisms overlap with the polygenic architectures of the disorders. We used single  
246 nucleotide polymorphism (SNP) data from the 20,170 adult healthy individuals with European  
247 ancestry available in UK Biobank. We estimated full and regional brain age for these individuals  
248 using 5-fold cross-validation in models trained on all healthy controls ( $n = 39,827$  aged 3-95  
249 years; 20,868 females, models trained per sex).

250           First, we performed one genome-wide association study (GWAS) per brain age gap using  
251 PLINK, including the first ten population components from multidimensional scaling, age, age ,  
252 sex, scanning site and Euler number as covariates. Next, we assessed heritability using LD score  
253 regression on the resulting summary statistics. In line with earlier results from twin studies<sup>13</sup>, our  
254 SNP-based analysis revealed significant heritability (**Fig. 3a**), with common SNPs explaining  
255 24% of the variance in brain age gap across all individuals (full brain,  $h^2_{\text{SNP}} = 0.24$ ,  $\text{SE} = 0.03$ )  
256 and 17-23% of the variance in regional brain age gaps (all  $\text{SE} < 0.03$ ).

257           Next, we assessed the overlap between the genetic underpinnings of brain age gap and  
258 common brain disorders. We gathered GWAS summary statistics for ASD, ADHD, SZ, BD, MS,  
259 major depression (MD), and Alzheimer's disease (AD) (see **online methods**). First, using LD  
260 score regression, we assessed the genetic correlation between these summary statistics and those  
261 from brain age gaps. Correlations were overall weak (**Suppl. Fig. 12**), with only one surviving  
262 FDR correction for the number of tests (cingulate brain age gap with ADHD). Lack of genetic  
263 correlation does not preclude genetic dependence as traits may have mixed effect directions  
264 across shared genetic variants<sup>14</sup>. Thus, we next used conjunctive FDR analyses to identify  
265 SNPs that are significantly associated with both brain age gap and disorders. We found  
266 significant independent loci showing pleiotropy between brain age gaps and all included

267 disorders (**Figure 3b**). Most loci were identified for SZ (2 occipital, 4 frontal, 3 temporal, 6  
268 parietal, 5 cingulate, 5 insula, 2 cerebellar/subcortical; 161 SNPs in total). Further, 5 independent  
269 loci for ASD (76 SNPs), 6 for ADHD (80 SNPs), 10 for BD (94 SNPs), 5 for MS (22 SNPs), 1  
270 for MD (14 SNPs), and 6 for AD (15 SNPs). **Suppl. Table 6** provides details. **Figure 3c** depicts  
271 the identified genes coloured by significance and sized by frequency. An intronic variant in  
272 protein coding gene *SATB2* at chromosome 2q33.1 was most frequently associated with brain age  
273 gaps and SZ. A missense variant in protein coding gene *SLC39A8* was associated with  
274 subcortical brain age gap and SZ and showed the strongest effect in all tested associations ( $P = 9$   
275  $\times 10^{-8}$ ).

276 Taken together, our results provide strong evidence that several common brain disorders  
277 are associated with an apparent aging of the brain, with effects observed at the full brain or  
278 regional level in SZ, PSYMIX, BD, MS, MDD, MCI and DEM; but not in ASD, ADHD or  
279 SZRISK. Importantly, our approach revealed differential neuroanatomical distribution of brain  
280 age gaps between several disorders. Associations with clinical and cognitive data in patients  
281 supported the functional relevance of the brain age gaps and genetic analyses in healthy  
282 individuals provided evidence that the brain age gaps are heritable, with overlapping genes  
283 between brain age gaps in healthy adults and common brain disorders.

284 Our approach of estimating regional brain age was useful to reveal differential spatial  
285 patterns between disorders. Whereas the implicated regions in the spatial brain age profiles of the  
286 disorders largely corresponded with previously reported structural abnormalities (e.g. frontal in  
287 SZ<sup>15</sup> and substantial subcortical volume loss in AD<sup>16</sup>), our regional brain age approach preserved  
288 the well-established benefit of down-sampling a large number of brain imaging features into a  
289 condensed and interpretable score without a total loss of spatial sensitivity. As such, the analysis  
290 revealed substantial differences in spatial aging profiles between disorders typically regarded as

291 neurodegenerative (MS, MCI, DEM) and neurodevelopmental, in particular SZ and PSYMIX.  
292 For example, whereas these disorders were all associated with increased brain age gap on the full  
293 brain level, regional analysis revealed interactions between the frontal brain age patterns  
294 observed in SZ and the cerebellar/subcortical patterns observed in MS and DEM, supporting  
295 spatial differences in apparent brain age. Moreover, significant associations with clinical and  
296 cognitive data, in particular with scores of the GAF and PANSS in SZ, with the EDSS in MS and  
297 with MMSE in the dementia spectrum demonstrated functional relevance of brain age gap  
298 beyond group differences. By gauging the dynamic associations between changes in brain age  
299 and clinical and cognitive function, future longitudinal studies may prove instrumental to dissect  
300 the large individual differences among patients with brain disorders, even within the same  
301 diagnostic category<sup>17</sup>. Furthermore, incorporating additional imaging modalities, voxel-level data  
302 or different segmentations at various levels of resolution will allow for estimation of tissue-  
303 specific brain age gaps or different regional gaps in future studies. Such approaches will also be  
304 useful to further investigate the apparent lack of brain age gap differences in ASD and ADHD. In  
305 contrast to research from other imaging phenotypes<sup>18,19</sup>, we did not observe case-control  
306 differences in brain age gaps for ASD or ADHD, nor group by age interactions (developmental  
307 delays might be reflected in a negative brain age gap in children). Brain age gaps based on  
308 different imaging modalities may capture different aspects of pathophysiology and will therefore  
309 yield an important contribution in future research.

310 Conceptually, brain age gaps reflect a prediction error from a machine learning model and  
311 can therefore be attributed to both noise (lack of model accuracy, insufficient data quality) and  
312 physiology (deviations from normal aging trajectories). The large training sample and accurate  
313 model performance, replication of results at different data quality criteria, as well as our  
314 approach of comparing brain age gaps of cases to a group of age-, sex- and scanner-matched

315 controls allowed us to reduce the impact of noise and to attribute variation in brain age gaps as  
316 likely related to biologically relevant differences. The physiological underpinnings of the brain  
317 age gaps are likely diverse, much like the polygenic nature of brain disorders and their  
318 profoundly heterogeneous symptomatology. They may reflect differences in disease severity,  
319 effects of comorbid disorders, substance use or other adverse lifestyle factors. Genetic analysis  
320 offers one way of exploring factors that influence phenotypic variation toward an improved  
321 understanding of the multi-faceted sources of lifespan trajectories in the brain. Here, we provided  
322 evidence that full and regional brain age gaps represent genetically influenced traits, and  
323 illustrated that the genetic variants associated with brain age gaps in healthy individuals partly  
324 overlap with those observed in ASD, ADHD, SZ, BD, MS, MD and AD. In line with  
325 accumulating evidence that common brain disorders are highly polygenic and partly  
326 overlapping<sup>20</sup> these results suggest shared molecular genetic mechanisms between brain age gaps  
327 and brain disorders. Statistical associations do not necessarily signify causation, and functional  
328 interpretations of the identified genes should be made with caution. Larger imaging genetics  
329 samples, in particular those including individuals with common brain disorders, may in the future  
330 allow the investigation of specificity of the implicated genes, and integrating a wider span of  
331 imaging modalities may increase both sensitivity and specificity.

332         In conclusion, we have established that the brain age gap is increased in several common  
333 brain disorders, sensitive to clinical and cognitive phenotypes and genetically influenced. Our  
334 results emphasize the potential of advanced lifespan modelling in the clinical neurosciences,  
335 highlighting the benefit of big data resources that cover a wide age span and conditions.  
336 Delineating dynamic lifespan trajectories within and across individuals will be essential to  
337 disentangle the pathophysiological complexity of brain disorders.

338 **Acknowledgements**

339 The author list between Ingrid Agartz and Mathias Zink is in alphabetic order. The authors were  
340 funded by the Research Council of Norway (276082 LifespanHealth (T.K.), 213837 (O.A.A.),  
341 223273 NORMENT (O.A.A.), 204966 (L.T.W.), 229129 (O.A.A.), 249795 (L.T.W.), 273345  
342 (L.T.W.), 283798 SYNSCHIZ (O.A.A.)), the South-Eastern Norway Regional Health Authority  
343 (2013-123 (O.A.A.), 2014-097 (L.T.W.), 2015-073 (L.T.W.), 2016083 (L.T.W.)), Stiftelsen  
344 Kristian Gerhard Jebsen, the European Research Council (ERC StG 802998 BRAINMINT  
345 (L.T.W.)), NVIDIA Corporation GPU Grant (T.K.), and the European Commission 7th  
346 Framework Programme (602450, IMAGEMEND (A.M.-L.)). The data used in this study were  
347 gathered from various sources. A detailed overview of the included cohorts and  
348 acknowledgement of their respective funding sources and cohort-specific details is provided in  
349 **Suppl. Table 1.** Data used in preparation of this article were obtained from the Alzheimer's  
350 Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu), from the AddNeuroMed  
351 consortium, and from the Pediatric Imaging, Neurocognition and Genetics Study (PING)  
352 database ([www.chd.ucsd.edu/research/ping-study.html](http://www.chd.ucsd.edu/research/ping-study.html), now shared through the NIMH Data  
353 Archive (NDA)). The investigators within the ADNI and PING contributed to the design and  
354 implementation of ADNI/PING and/or provided data but did not participate in analysis or writing  
355 of this report. This publication is solely the responsibility of the authors and does not necessarily  
356 represent the views of the National Institutes of Health or PING investigators. Complete listings  
357 of participating sites and study investigators can be found at [http://adni.loni.usc.edu/wp-](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf)  
358 [content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf) and [https://ping-](https://ping-dataportal.ucsd.edu/sharing/Authors10222012.pdf)  
359 [dataportal.ucsd.edu/sharing/Authors10222012.pdf](https://ping-dataportal.ucsd.edu/sharing/Authors10222012.pdf). The AddNeuroMed consortium was led by  
360 Simon Lovestone, Bruno Vellas, Patrizia Mecocci, Magda Tsolaki, Iwona Kłoszewska, Hilikka  
361 Soininen.

362

363 **Author contributions**

364 T.K. and L.T.W. conceived the study; T.K., N.T.D. and L.T.W. pre-processed all data in  
365 Freesurfer; N.T.D., M.J.L., C.L.B, L.B.N., L.T.W. and T.K. performed quality control of the  
366 data; T.K. performed the analysis with contributions from L.T.W. and D.v.d.M.; T.K., L.T.W.,  
367 N.T.D., D.v.d.M. and O.A.A. contributed to interpretation of the results. All remaining authors  
368 were involved in data collection at various sites as well as cohort-specific tasks. T.K. and L.T.W.  
369 wrote the first draft of the paper and all authors contributed to and approved the final manuscript.

370 **Competing financial interests**

371 Some authors received educational speaker's honorarium from Lundbeck (O.A. Andreassen, A.  
372 Bertolino, T. Elvsåshagen, M. Zink, N. I. Landrø), Sunovion (O.A. Andreassen), Shire (B.  
373 Franke), Medice (B. Franke), Otsuka (A. Bertolino, M. Zink) and Janssen (A. Bertolino), Roche  
374 (M. Zink), Ferrer (M. Zink), Trommsdorff (M. Zink), Servier (M. Zink), all of these unrelated to  
375 this work. A. Bertolino is a stockholder of Hoffmann-La Roche Ltd and has received consultant  
376 fees from Biogen Idec. E. G. Celius and H. F. Harbo have received travel support, honoraria for  
377 advice and lecturing from Almirall (Celius), Biogen Idec (both), Genzyme (both), Merck (both),  
378 Novartis(both), Roche (both), Sanofi-Aventis (both) and Teva (both). They have received  
379 unrestricted research grants from Novartis (Celius, Harbo), Biogen Idec (Celius) and Genzyme  
380 (Celius). G. Pergola has been the academic supervisor of a Roche collaboration grant (years  
381 2015-16) that funds his salary. None of the mentioned external parties had any role in the  
382 analysis, writing or decision to publish this work. Other authors declare no competing financial  
383 interests.

384

385 **Members of the Karolinska Schizophrenia Project (KaSP)**

386 Farde L<sup>4</sup>, Flyckt L<sup>4</sup>, Engberg G<sup>70</sup>, Erhardt S<sup>70</sup>, Fatouros-Bergman H<sup>4</sup>, Cervenka S<sup>4</sup>, Schwieler  
387 L<sup>70</sup>, Piehl F<sup>71</sup>, Agartz I<sup>1,3,4</sup>, Collste K<sup>4</sup>, Victorsson P<sup>4</sup>, Malmqvist A<sup>70</sup>, Hedberg M<sup>70</sup>, Orhan F<sup>70</sup>

388 <sup>70</sup>Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden

389 <sup>71</sup>Neuroimmunology Unit, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm,

390 Sweden

391 **References**

- 392 1 WHO. *World Health Statistics 2016*. (2016).  
393 2 Insel, T. R. & Cuthbert, B. N. Brain disorders? Precisely. *Science* **348**, 499-500,  
394 doi:10.1126/science.aab2358 (2015).  
395 3 Prince, M. *et al.* No health without mental health. *Lancet* **370**, 859-877,  
396 doi:10.1016/S0140-6736(07)61238-0 (2007).  
397 4 Parikshak, N. N., Gandal, M. J. & Geschwind, D. H. Systems biology and gene networks  
398 in neurodevelopmental and neurodegenerative disorders. *Nat Rev Genet* **16**, 441-458,  
399 doi:10.1038/nrg3934 (2015).  
400 5 Marin, O. Developmental timing and critical windows for the treatment of psychiatric  
401 disorders. *Nat Med* **22**, 1229-1238, doi:10.1038/nm.4225 (2016).  
402 6 Insel, T. R. Rethinking schizophrenia. *Nature* **468**, 187-193, doi:Doi  
403 10.1038/Nature09552 (2010).  
404 7 Aubert-Broche, B. *et al.* Onset of multiple sclerosis before adulthood leads to failure of  
405 age-expected brain growth. *Neurology* **83**, 2140-2146,  
406 doi:10.1212/WNL.0000000000001045 (2014).  
407 8 Masters, C. L. *et al.* Alzheimer's disease. *Nat Rev Dis Primers* **1**, 15056,  
408 doi:10.1038/nrdp.2015.56 (2015).  
409 9 Dosenbach, N. U. *et al.* Prediction of individual brain maturity using fMRI. *Science* **329**,  
410 1358-1361, doi:10.1126/science.1194144 (2010).  
411 10 Franke, K., Ziegler, G., Kloppel, S., Gaser, C. & Alzheimer's Disease Neuroimaging, I.  
412 Estimating the age of healthy subjects from T1-weighted MRI scans using kernel  
413 methods: exploring the influence of various parameters. *Neuroimage* **50**, 883-892,  
414 doi:10.1016/j.neuroimage.2010.01.005 (2010).  
415 11 Cole, J. H. & Franke, K. Predicting Age Using Neuroimaging: Innovative Brain Ageing  
416 Biomarkers. *Trends Neurosci* **40**, 681-690, doi:10.1016/j.tins.2017.10.001 (2017).  
417 12 Ritchie, S. J. *et al.* Sex Differences in the Adult Human Brain: Evidence from 5216 UK  
418 Biobank Participants. *Cereb Cortex* **28**, 2959-2975, doi:10.1093/cercor/bhy109 (2018).  
419 13 Cole, J. H. *et al.* Predicting brain age with deep learning from raw imaging data results in  
420 a reliable and heritable biomarker. *Neuroimage* **163**, 115-124,  
421 doi:10.1016/j.neuroimage.2017.07.059 (2017).  
422 14 Bansal, V. *et al.* Genome-wide association study results for educational attainment aid in  
423 identifying genetic heterogeneity of schizophrenia. *Nature Communications* **9**, 3078,  
424 doi:10.1038/s41467-018-05510-z (2018).



- 425 15 Ellison-Wright, I. & Bullmore, E. Anatomy of bipolar disorder and schizophrenia: a meta-  
426 analysis. *Schizophrenia research* **117**, 1-12, doi:10.1016/j.schres.2009.12.022 (2010).
- 427 16 Jernigan, T. L., Salmon, D. P., Butters, N. & Hesselink, J. R. Cerebral structure on MRI,  
428 Part II: Specific changes in Alzheimer's and Huntington's diseases. *Biological psychiatry*  
429 **29**, 68-81 (1991).
- 430 17 Wolfers, T. *et al.* Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar  
431 Disorder Using Normative Models. *Jama Psychiat* **75**, 1146-1155,  
432 doi:10.1001/jamapsychiatry.2018.2467 (2018).
- 433 18 Ecker, C., Bookheimer, S. Y. & Murphy, D. G. Neuroimaging in autism spectrum  
434 disorder: brain structure and function across the lifespan. *Lancet Neurol* **14**, 1121-1134,  
435 doi:10.1016/S1474-4422(15)00050-2 (2015).
- 436 19 Faraone, S. V. *et al.* Attention-deficit/hyperactivity disorder. *Nature Reviews Disease*  
437 *Primers* **1**, 15020, doi:10.1038/nrdp.2015.20 (2015).
- 438 20 Andreassen, O. A. *et al.* Genetic pleiotropy between multiple sclerosis and schizophrenia  
439 but not bipolar disorder: differential involvement of immune-related gene loci. *Molecular*  
440 *psychiatry* **20**, 207 (2015).

441  
442

#### 443 **Figure legends**

444

445 **Figure 1: Sample distributions and imaging features used for brain age prediction. a,** Age

446 distributions of the training (left) and the ten test samples (right) per sex and diagnosis. The grey

447 shades behind each clinical group reflect its age-, sex- and site-matched control group. **b,** Cortical

448 features from the Human Connectome Project (HCP) atlas as well as cerebellar/subcortical

449 features used for brain age prediction. Colours were assigned randomly to each feature. All

450 features were used in the full brain feature set (left), whereas only those from specific regions

451 (occipital, frontal, temporal, parietal, cingulate, insula, cerebellar/subcortical) were included in

452 the regional feature set (right). For illustration purpose, the left hemisphere is shown.

453

454 **Figure 2: Apparent brain aging is common in several brain disorders and sensitive to**

455 **clinical and cognitive measures. a,** The gap between chronological age and brain age was

456 increased in several disorders. The grey shades behind each clinical group reflect its age-, sex-

457 and site-matched controls. The test samples comprised n=925 ASD / n=925 HC, n=725 ADHD /

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

458 n=725 HC, n=94 SZRISK / n=94 HC, n=1110 SZ / n=1110 HC, n=300 PSYMIX / n=300 HC,  
459 n=459 BD / n=459 HC, n=254 MS / n=254 HC, n=208 MDD / n=208 HC, n=974 MCI / n=974  
460 HC, n=739 DEM / n=739 HC; in total n=10,141 independent subjects. Cohen's d effect sizes  
461 (pooled standard deviation units) and two-sided P-values are provided. **b**, Several disorders  
462 showed specific patterns in regional brain age gaps. Colours indicate Cohen's d effect sizes for  
463 group comparisons. Sample size as specified in panel a. Corresponding correlation matrix of the  
464 effect sizes is depicted in **Suppl. Fig. 9**. **c**, Effect sizes of significant region by group interactions  
465 from repeated measures ANOVAs run for each combination of regions and groups (1260 tests in  
466 total). Sample size as specified in panel a yet excluding HC; n=5788 independent subjects. Only  
467 significant ( $p < \text{FDR}$ ; Benjamini-Hochberg) effects are shown. **Suppl. Fig. 10** depicts effect sizes  
468 for all 1260 tests. **d**, Correlation coefficients for linear associations between brain age gaps and  
469 cognitive and clinical scores. Sample size comprised n=389 SZ for  $\text{GAF}_{\text{symptom}}$ , n=269 SZ for  
470  $\text{GAF}_{\text{function}}$ , n=646 SZ for  $\text{PANSS}_{\text{positive}}$ , n=626 SZ for  $\text{PANSS}_{\text{negative}}$ , n=195 MS for EDSS, n=907  
471 MCI and n=686 DEM for MMSE. Associations were computed using linear models accounting  
472 for age, age<sup>2</sup>, sex, scanning site and Euler number, and the resulting t-statistics were transformed  
473 to r. Significant ( $P < \text{FDR}$ ; Benjamini-Hochberg; two-sided) associations are marked with a black  
474 box. Corresponding scatter plots are depicted in **Suppl. Fig 11**.

475  
476 **Figure 3: The brain age gaps are heritable, and the genetic underpinnings overlap with**  
477 **those observed for several disorders.** Genetic analyses were performed using data from  
478 n=20,170 healthy adult individuals with European ancestry **a**, Heritability ( $h^2$ ) estimated using  
479 LD Score regression. Error bars reflect standard error. **b**, Significantly ( $P < \text{FDR}$ ) overlapping loci  
480 between brain age gaps and disorders, identified using *conjunctional FDR*. **c**, Corresponding to

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

481 panel b, the overlapping genes across all disorders, coloured by significance and sized by  
482 frequency of detection.

483 **Online methods**

484 Additional information is available in the *Life Sciences Reporting Summary*.

485 *Samples*

486 We have included data collected through collaborations, data sharing platforms, consortia as well  
487 as available in-house cohorts. No statistical methods were used to pre-determine sample sizes.  
488 We included as much data as we could gather (brain scans from N=45,615 individuals) and  
489 sample size of individual clinical groups is thus based on data availability. **Suppl. Tables 1 - 3**  
490 provide detailed information on the individual cohorts. All included cohorts have been published  
491 on, and we refer to a list of publications that can be consulted for a more detailed overview of  
492 cohort characteristics. Data collection in each cohort was performed with participants' written  
493 informed consent and with approval by the respective local Institutional Review Boards.

494 *Image pre-processing and quality control*

495 Raw T1 data for all study participants were stored and analysed locally at University of Oslo,  
496 following a harmonized analysis protocol applied to each individual subject data (**Suppl. Fig. 1**).  
497 We performed automated surface-based morphometry and subcortical segmentation using  
498 Freesurfer 5.3<sup>21</sup>. We deployed an automated quality control protocol executed within each of the  
499 contributing cohorts that excluded potential outliers based on the Euler number<sup>22</sup> of the respective  
500 Freesurfer segmentations. Euler number captures the topological complexity of the uncorrected  
501 Freesurfer surfaces and thus comprises a proxy of data quality<sup>22</sup>. In brief, for each scanning site  
502 we regressed age, age<sup>2</sup> and sex from the Euler number of the left and right hemispheres and  
503 identified scans that exceeded 3 standard deviations (SD) on either of the residualized Euler  
504 numbers. **Suppl. Fig. 13** provides a validation of the approach against manual quality control.  
505 Data from a total of 977 individuals was excluded in this step, yielding 45,615 subjects for the

506 main analysis. To further minimize confounding effects of data quality<sup>23</sup>, we performed  
507 supplementary analyses using a subset of data, where a more stringent threshold was used for  
508 exclusion (1 SD on Euler numbers). Thus, supplemental analysis provides a sanity check with  
509 those subjects excluded (sample size:  $n = 40,301$ ).

### 510 *Brain age prediction*

511 We utilized a recent multi-modal cortical parcellation scheme<sup>24</sup> to extract cortical thickness, area  
512 and volume for 180 regions of interest (ROI) per hemisphere. In addition, we extracted the classic  
513 set of cerebellar/subcortical and cortical summary statistics<sup>21</sup>. This yielded a total set of 1118  
514 structural brain imaging features (360/360/360/38 for cortical thickness/area/volume as well as  
515 cerebellar/subcortical and cortical summary statistics, respectively).

516 We used machine learning on this feature set to predict the age of each individual's brain.  
517 First, we split the available data into a training sample and ten independent test samples (**Fig. 1a**).  
518 The test samples in total comprised 5788 individuals with brain disorders and 4353 healthy  
519 controls. For each of the ten clinical groups, we selected a set of healthy controls from the pool of  
520 4353 individuals, matched for age, sex and scanning site using propensity score matching<sup>25</sup>.  
521 Thus, data from some healthy individuals acted as control data in several test samples, yet each  
522 test sample had the same number of patients and controls and all subjects in the test samples were  
523 independent of the subjects in the training sample. The remaining datasets (45,615 –  
524 (5788+4353) = 35,474) went into the training set. For each sex, we trained machine learning  
525 models based on gradient tree boosting<sup>26</sup> utilizing the *xgboost* package in R<sup>27</sup>, chosen due to its  
526 resource efficiency and demonstrated superior performance in previous machine learning  
527 competitions<sup>26</sup>, to predict the age of the brain using data available in the training set. First, model  
528 parameters were tuned using a 5-fold cross-validation of the training data. This step identified the

529 optimal number of model training iterations by assessing the prediction error for 1500 rounds and  
530 implementing an early stopping if the performance did not improve for 20 rounds. Based on  
531 previous experience, the learning rate was pre-set to  $\eta=0.01$  and all other parameters were set to  
532 default<sup>27</sup> for linear *xgboost* tree models. After determining the optimal number of training  
533 iterations, the full set of training data was used to train the final models with the adjusted *nrounds*  
534 parameter. These models were used to predict brain age in the test samples, and the brain age gap  
535 (deviation between brain and chronological age) was computed. In line with a recent  
536 recommendation<sup>28</sup>, all statistical analyses on the brain age gap accounted for age, sex,  
537 scanning site and Euler number. In addition, to assess overall model performance, prediction  
538 models were cross-validated within the training set using a 5-fold cross validation, each fold  
539 implementing the above described training procedure and testing on the hold-out part of the  
540 training set. Brain age predictions on the level of individual brain regions followed the same  
541 procedures as those described for the full brain level, except that the feature set was reduced to  
542 cover only those features that overlapped more than 50% with a given lobe. Regions were  
543 defined following the Freesurfer *lobesStrict* segmentation as *occipital*, *frontal*, *temporal*, *parietal*,  
544 *cingulate* and *insula*. In addition, given the limited number of cerebellar features available in the  
545 Freesurfer summary statistics, cerebellar and subcortical features were grouped into a  
546 *cerebellar/subcortical* region (**Fig. 1b**). For additional validation, we compared our *xgboost*  
547 approach against two other approaches (**Suppl. Fig. 3**). One approach implemented a different  
548 machine learning algorithm on the same set of features (*slm* from the *care* package<sup>29</sup>), whereas  
549 the other approach made use of a fully independent processing pipeline, feature set and algorithm  
550 ([github.com/james-cole/brainageR](https://github.com/james-cole/brainageR)<sup>13,30</sup>). Furthermore, we assessed the impact of sample size on  
551 model performance by creating random subsets of data with sample sizes of 100, 500, 1000,

Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

552 2000, 5000, 10,000, and 20,000 individuals (40 random subsets per sample size). For each subset  
553 and sample size we assessed model performance using cross-validation (**Suppl. Fig. 5**).

554 The genetic analysis was performed in UK Biobank data, which was part of the training  
555 set in the main analysis. We thus trained different brain age models for the genetic analysis. We  
556 selected all healthy subjects and estimated their brain age using a 5-fold cross-validation  
557 approach, like the one performed when validating performance of the training set. The resulting  
558 unbiased estimates of brain age gaps for all UK Biobank individuals with genetic data available  
559 went into the genome-wide association analysis, LD score regression and conjunctive FDR.

#### 560 *Main statistical analysis framework*

561 We performed both mega- (across cohorts) and meta- (within cohort) analyses. To estimate group  
562 effects on a given measure in a mega-analysis framework, we computed the effect of diagnosis in  
563 relation to the healthy controls for each of the ten test samples in a linear model accounting for  
564 age, age , sex, scanning site and Euler number. Cohen's d effect sizes were estimated based on  
565 contrast t-statistics<sup>31</sup> following **Formula 1**:

$$d = \frac{t(n_1 + n_2)}{\sqrt{n_1 n_2} \sqrt{df}} \quad (1)$$

566 For the meta-analysis, similar models were computed within cohorts. In addition to estimating  
567 Cohen's d (**Formula 1**), we estimated the variance of d following **Formula 2**.

$$v = \left( \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2 - 2)} \right) \left( \frac{n_1 + n_2}{n_1 + n_2 - 2} \right) \quad (2)$$

568 Cumulative effects across cohorts were then estimated using a variance-weighted random-effects  
569 model as implemented in the *metafor* package in R<sup>32</sup>.

570 Data distributions were assumed to be normal, but this was not formally tested. Data collection  
571 and analysis were not performed blind to the conditions of the experiments.

572 *Assessment of regional specificity*

573 In **Suppl. Fig 9**, we performed clustering of effect sizes from Figure 2b using heatmap.2 from the  
574 *gplots* package<sup>33</sup> in R. A Spearman correlation matrix was computed based on the case-control  
575 effect sizes obtained from each test sample and region and hierarchical clustering was performed  
576 using the default settings. To further explore regional specificity, we performed an analysis that  
577 involved only the clinical groups. We regressed age, age , sex, scanning site and Euler number  
578 from the brain age gaps in each test sample. Next, we joined data from each pair of clinical  
579 groups and each pair of regions for repeated measures analysis of variance and estimated the  
580 effect sizes of region x group interactions (1260 ANOVAs in total). The significant interaction  
581 effects were visualized in **Figure 2c** using the *circlize* package<sup>34</sup> in R.

582 *Genetic analyses*

583 We restricted all genetic analyses to individuals from the UK Biobank with European ancestry, as  
584 determined by the UK Biobank study team<sup>35</sup>. We applied standard quality control procedures to  
585 the UK Biobank v3 imputed genetic data. In brief, we removed SNPs with an imputation quality  
586 score below 0.5, with a minor allele frequency less than .05, missing in more than 5% of  
587 individuals, and failing the Hardy Weinberg equilibrium tests at a  $p < 1 \times 10^{-6}$ , yielding SNP data  
588 from 20,170 adult healthy individuals. We performed a genome-wide association analysis using  
589 PLINK v1.9<sup>36</sup>, accounting the analysis for 10 genetic principal components, age, age , sex,  
590 scanning site and Euler number. We used LD Score regression<sup>37</sup> to estimate narrow sense  
591 heritability.

592 Furthermore, we used cross-trait LD Score regression<sup>37,38</sup> to calculate genetic correlations,  
593 and conjunctive FDR analyses<sup>39,40</sup> to assess genetic overlap between two complex traits. We  
594 gathered genome-wide association analysis (GWAS) summary statistics for ASD<sup>41</sup>, ADHD<sup>42</sup>,



Kaufmann et al., Genetics of brain age suggest an overlap with common brain disorders

595 SZ<sup>43</sup>, BD<sup>44</sup>, MS<sup>45</sup>, MD<sup>46</sup>, and AD<sup>47</sup>; and assessed genetic overlap with brain age gap genetics.  
596 The MHC region was excluded from all analysis. Conjunctive FDR was run for each pair of  
597 full brain / regional brain age gap and group, using conjunctive FDR threshold of 0.05. SNPs  
598 were annotated using the Ensembl Variant Effect Predictor<sup>48</sup>.

### 599 *Cognitive and clinical associations*

600 Cognitive and clinical associations were tested in subsets based on data availability and were  
601 performed in clinical groups only (excluding controls) as described in the main text. Using linear  
602 models accounting for age, age<sup>2</sup>, sex, scanning site and Euler number we associated brain age  
603 gaps with scores of the Global Assessment of Functioning scale<sup>49</sup> (GAF), the Positive and  
604 Negative Syndrome Scale<sup>50</sup> (PANSS), the Expanded Disability Status Scale<sup>51</sup> (EDSS) and Mini  
605 Mental State Examination scores<sup>52</sup> (MMSE). The t-statistics of the linear models were  
606 transformed to r, thus the correlation coefficients depicted in Fig 2d essentially reflect a partial  
607 correlation between full brain / regional brain age gaps and clinical/cognitive scores, controlling  
608 for confounding effects of age, sex, site and image quality.

### 609 **Code availability.**

610 Code needed to run brain age prediction models is available at [github.com/tobias-kaufmann](https://github.com/tobias-kaufmann) (see  
611 Data availability). Additional R statistics<sup>53</sup> code is available from the authors upon request.

### 612 **Data availability**

613 The raw data incorporated in this work were gathered from various resources. Material requests  
614 will need to be placed with individual PIs. A detailed overview of the included cohorts is  
615 provided in **Suppl. Table 1**. GWAS summary statistics for the brain age gaps as well as the  
616 models needed to predict brain age in independent cohorts are available at [github.com/tobias-](https://github.com/tobias-kaufmann)  
617 [kaufmann](https://github.com/tobias-kaufmann).

618 **Methods-only References**

- 619 21 Fischl, B. *et al.* Whole brain segmentation: Automated labeling of neuroanatomical  
620 structures in the human brain. *Neuron* **33**, 341-355, doi:Doi 10.1016/S0896-  
621 6273(02)00569-X (2002).
- 622 22 Rosen, A. F. G. *et al.* Quantitative assessment of structural image quality. *Neuroimage*  
623 **169**, 407-418, doi:10.1016/j.neuroimage.2017.12.059 (2018).
- 624 23 Smith, S. M. & Nichols, T. E. Statistical Challenges in "Big Data" Human Neuroimaging.  
625 *Neuron* **97**, 263-268, doi:10.1016/j.neuron.2017.12.018 (2018).
- 626 24 Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**,  
627 171-178, doi:10.1038/nature18933 (2016).
- 628 25 Ho, D., Imai, K., King, G. & Stuart, E. A. MatchIt: Nonparametric Preprocessing for  
629 Parametric Causal Inference. *Journal of Statistical Software* **42**, 1-28 (2011).
- 630 26 Chen, T. & Guestrin, C. in *Proceedings of the 22nd ACM SIGKDD International*  
631 *Conference on Knowledge Discovery and Data Mining* 785-794 (ACM, San Francisco,  
632 California, USA, 2016).
- 633 27 Tianqi, C., Tong, H., Benesty, M., Khotilovich, V. & Tang, Y. Xgboost: extreme gradient  
634 boosting. R package v0.4-2. (2015).
- 635 28 Le, T. T. *et al.* A Nonlinear Simulation Framework Supports Adjusting for Age When  
636 Analyzing BrainAGE. *Front Aging Neurosci* **10**, 317, doi:10.3389/fnagi.2018.00317  
637 (2018).
- 638 29 Zuber, V. & Strimmer, K. Care. R package v 1.1.10. *Care. R package v 1.1.10* (2017).
- 639 30 Cole, J. H. *et al.* Brain age predicts mortality. *Molecular psychiatry* **23**, 1385-1392,  
640 doi:10.1038/mp.2017.62 (2018).
- 641 31 Nakagawa, S. & Cuthill, I. C. Effect size, confidence interval and statistical significance:  
642 a practical guide for biologists. *Biol Rev Camb Philos Soc* **82**, 591-605,  
643 doi:10.1111/j.1469-185X.2007.00027.x (2007).
- 644 32 Viechtbauer, W. Conducting meta-analysis in R with the metafor package. *Journal of*  
645 *Statistical Software* **36**, 1-48 (2010).
- 646 33 Warnes, G. R. *et al.* R Package gplots: Various R Programming Tools for Plotting Data.  
647 (2016).
- 648 34 Gu, Z. R Package circlize: Circular Visualization. (2017).
- 649 35 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.  
650 *Nature* **562**, 203-209, doi:10.1038/s41586-018-0579-z (2018).
- 651 36 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based  
652 linkage analyses. *American Journal of Human Genetics* **81**, 559-575, doi:10.1086/519795  
653 (2007).
- 654 37 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from  
655 polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295,  
656 doi:10.1038/ng.3211 (2015).
- 657 38 Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits.  
658 *Nature genetics* **47**, 1236-1241, doi:10.1038/ng.3406 (2015).
- 659 39 Nichols, T., Brett, M., Andersson, J., Wager, T. & Poline, J.-B. Valid conjunction  
660 inference with the minimum statistic. *Neuroimage* **25**, 653-660 (2005).
- 661 40 Andreassen, O. A. *et al.* Improved detection of common variants associated with  
662 schizophrenia by leveraging pleiotropy with cardiovascular-disease risk factors. *The*  
663 *American Journal of Human Genetics* **92**, 197-209 (2013).

- 664 41 Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum  
665 disorder. *Nature genetics* **51**, 431-444, doi:10.1038/s41588-019-0344-8 (2019).
- 666 42 Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention  
667 deficit/hyperactivity disorder. *Nature genetics*, doi:10.1038/s41588-018-0269-7 (2018).
- 668 43 Schizophrenia Working Group of the PGC *et al.* Biological insights from 108  
669 schizophrenia-associated genetic loci. *Nature* **511**, 421, doi:10.1038/nature13595 (2014).
- 670 44 Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with  
671 bipolar disorder. *Nature genetics* **51**, 793-803, doi:10.1038/s41588-019-0397-8 (2019).
- 672 45 Patsopoulos, N. *et al.* The Multiple Sclerosis Genomic Map: Role of peripheral immune  
673 cells and resident microglia in susceptibility. *bioRxiv*, 143933, doi:10.1101/143933  
674 (2017).
- 675 46 Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine  
676 the genetic architecture of major depression. *Nature genetics* **50**, 668-681,  
677 doi:10.1038/s41588-018-0090-3 (2018).
- 678 47 Lambert, J.-C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility  
679 loci for Alzheimer's disease. *Nature genetics* **45**, 1452 (2013).
- 680 48 McLaren, W. *et al.* The ensembl variant effect predictor. *Genome biology* **17**, 122 (2016).
- 681 49 Pedersen, G. & Karterud, S. The symptom and function dimensions of the Global  
682 Assessment of Functioning (GAF) scale. *Comprehensive Psychiatry* **53**, 292-298 (2012).
- 683 50 Kay, S. R., Fiszbein, A. & Opfer, L. A. The positive and negative syndrome scale  
684 (PANSS) for schizophrenia. *Schizophrenia Bull* **13**, 261 (1987).
- 685 51 Kurtzke, J. F. Rating neurologic impairment in multiple sclerosis: an expanded disability  
686 status scale (EDSS). *Neurology* **33**, 1444-1444 (1983).
- 687 52 Folstein, M. F., Folstein, S. E. & McHugh, P. R. "Mini-mental state": a practical method  
688 for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*  
689 **12**, 189-198 (1975).
- 690 53 R Core Team. R: A language and environment for statistical computing. *R Foundation for*  
691 *Statistical Computing, Vienna, Austria.* (2013).
- 692